Semantic Terminology Services: Experiences from the FACET Project

Doug Tudhope
Hypermedia Research Unit
University of Glamorgan

DELOS Workshop, Lund, June 2004

Presentation

- NKOS overview
- FACET Project
  - Semantic expansion
  - Standalone system
  - Qualitative evaluation
  - Web Demonstrator
- Lessons learned
  - Need for standards
- Future work
  - Semantic expansion service
- References

Taxonomy of Knowledge Organisation Systems

- Term Lists
  - Authority Files, Glossaries, Gazetteers, Dictionaries
- Classification and Categorization
  - Subject Headings
  - Classification Schemes and Taxonomies
    - eg DDC, scientific taxonomies
- Relationship Schemes
  - Thesauri
  - Semantic Networks (eg WordNet)
  - (Ontologies)

Hodg00, http://www.clir.org/pubs/abstract/pub91abst.html

Semantic KOS

Thesauri

- 3 Standard Relationships between concepts (Aitc00)
  - Equivalence, Hierarchical, Associative
- Inherent domain lexicon (lead-in vocabulary)
- Concept definitions and warrant (Scope Notes)

Ontologies

- Higher level conceptualisation (McGu02, Noy)
  - formal definition of relationships
  - inference rules and definition of roles (sometimes)

KOS an element of ontologies and schemas

Jaco03, Ontologies and the Semantic Web,.
ASIST Bulletin, April/May 2003, *Special Issue on Semantic Web*

Terminology Services from Koch04 Structured Overview - Activities to advance the powerful use of vocabularies

Searching for concepts

- schemes in registries
- concepts/terms in taxonomy servers

Search support for queries

- collection finding
- cross-searching, cross-browsing, mapping services
- KOS browsing and user interface/visualisation
- query expansion, disambiguation
- automatic indexing and classification
- extraction/mining of terms
- translation support using vocabularies

FACET: faceted knowledge organisation for semantic retrieval

University of Glamorgan, Science Museum

KOS creation and maintenance

Mapping, merging vocabularies

Document creation and maintenance

Indexing, classification, annotation

intellectual, automatic

faceted, multi-concept bestmatch search

semantic expansion as browsing service

faceted thesaurus search interface

standalone and Web demonstrators

Discovery of services and databases/collections

Searching for concepts --> controlled terminology, auto-disambiguation

Querying and result display

Cross-searching, cross-browsing, mapping services

KOS browsing and user interface/visualisation

Query expansion

Extraction/mining of terms

Translation support using vocabularies

Content integration and mediation

Recent Sources

NKOS: Networked Knowledge Organization Systems/Services

- http://jodi.ecs.soton.ac.uk/?vol=4&iss=4 *NKOS JoDI Special Issue*
- http://www.multites.com/conference03.htm *MultiTes Conference*
- http://nkos.slis.kent.edu/ *JCDL and ECDL Workshops 2003*
- http://www.lub.lu.se/SEMKOS/ *SEMKOS IP Proposal Resources*
- http://www.digicult.info *Cultural Heritage review site*

Semantic Web - RDF/XML, RDF Schema, Metalog, OWL

- http://www.w3.org/2001/sw/ *W3C Semantic Web Activity*
- http://www.semanticweb.org/
- http://ontoweb.aifb.uni-karlsruhe.de/ *OntoWeb*
- http://www.w3c.rl.ac.uk/SWAD/thesaurus.html *SWAD-Europe Thesaurus index*

Semantic Grid - Semantic Web, Web service, eScience, GRID links

- http://www.semanticgrid.org/
- http://www.w3.org/2002/ws/ *W3C Web Services Activity*
- http://www.ariadne.ac.uk/issue29/gardner/intro.html *Gardner's Intro to Web Services*
- http://www.ukoln.ac.uk/events/jisc-terminology/ *JISC Terminology Services Workshop*

FACET - Faceted Access to Cultural hEritage Terminology

FACET - a collaborative project investigating the potential of semantic term expansion in retrieval

Aims:

- Integration of thesaurus into the interface
- Semantic term expansion and matching function taking advantage of facet structure

http://www.comp.glam.ac.uk/~FACET/

FACET Collaborators

- Research Council Funding: EPSRC 3 years
- National Museum of Science and Industry (NMSI):
  National Railway Museum and Science Museum Collections Database
- J. Paul Getty Trust
  Art and Architecture Thesaurus (AAT)
- Museum Documentation Association (MDA)
  Railway Thesaurus
- Canadian Heritage Information Network (CHIN)
  Advisors

The Thesaurus in Retrieval

a) as a ***search thesaurus*** with a (web) ***free text*** search engine
resource for query refinement (interactive or automatic)

b) searching and indexing with ***controlled vocabulary*** indexed datasets
- immediate application area of FACET project

In searching, thesaurus relationships conventionally used to
expand synonyms
include narrower terms in a query

-- but can also be used in more general **semantic term expansion**

Semantic Term Expansion

Reasoning over thesaurus semantic relationships
allows the system to play an active role

- Ranking of matching items in a result set
- Automatic suggestion of terms to be considered for query
- Query reformulation and ‘more like this’ option
- Augmented Browsing tools – semantic expansion

Underpinning technologies:

- Measures of distance over the semantic index space
- Matching Function for sets of terms

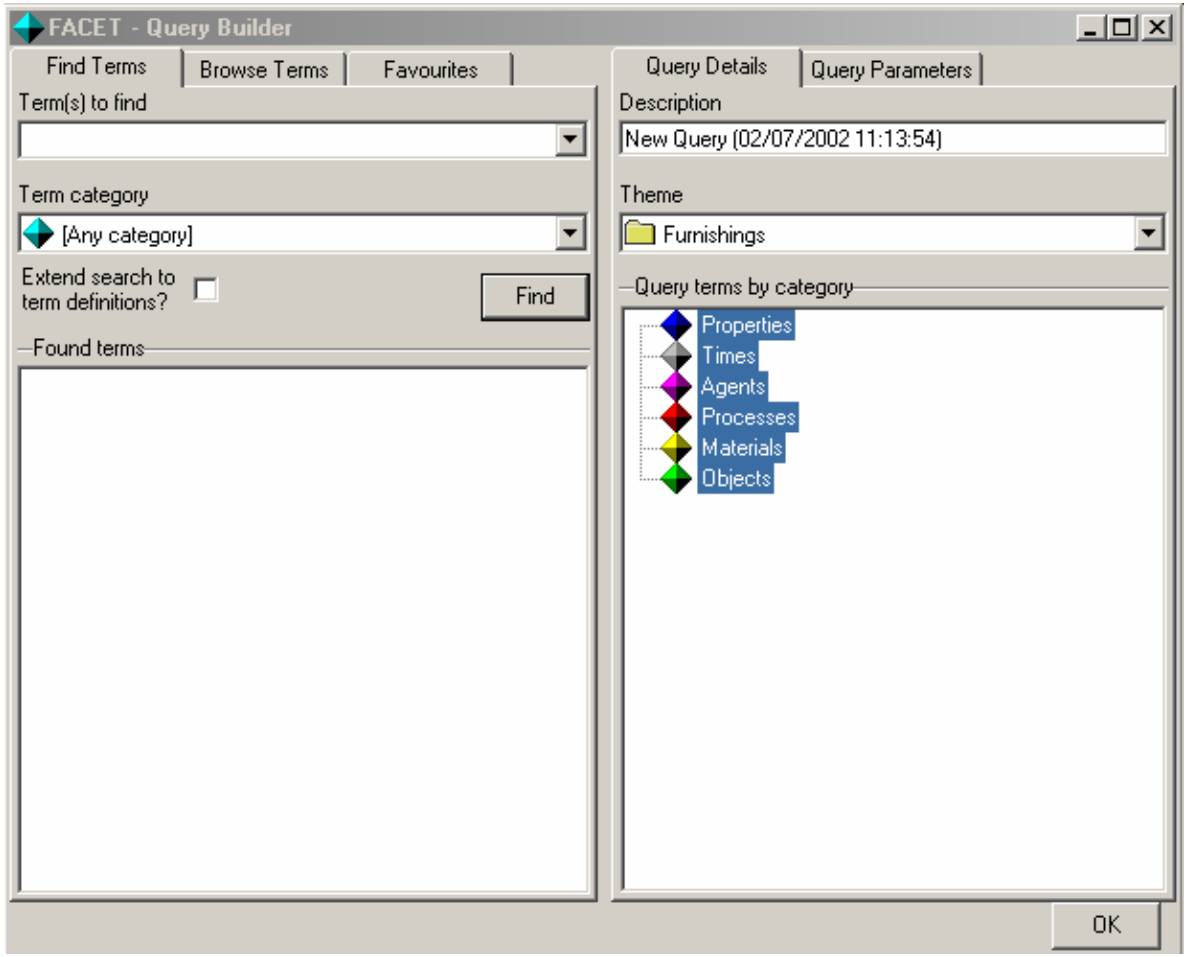
FACET Prototype

- SQLServer database: collections DB and Thesaurus

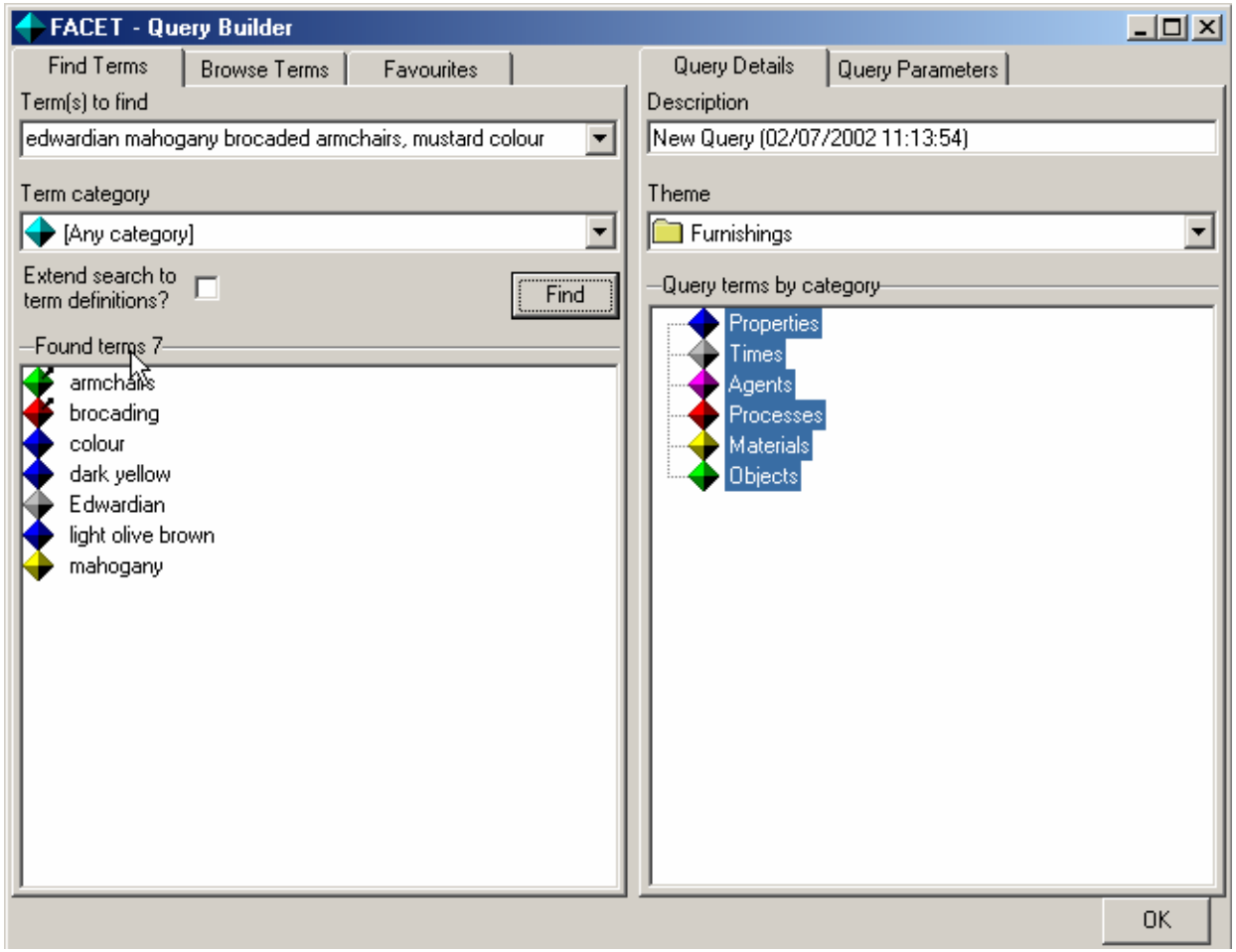
- C++ thesaurus term expansion engine
- Dual thesaurus representations
  - database
  - in-memory data structure

- Visual Basic and Web client interfaces
  - ‘Find Term’ mapping to terms, alternates, scope notes
  - Browse hierarchies
  - Semantic browsing
  - Query Builder
  - Ranked results

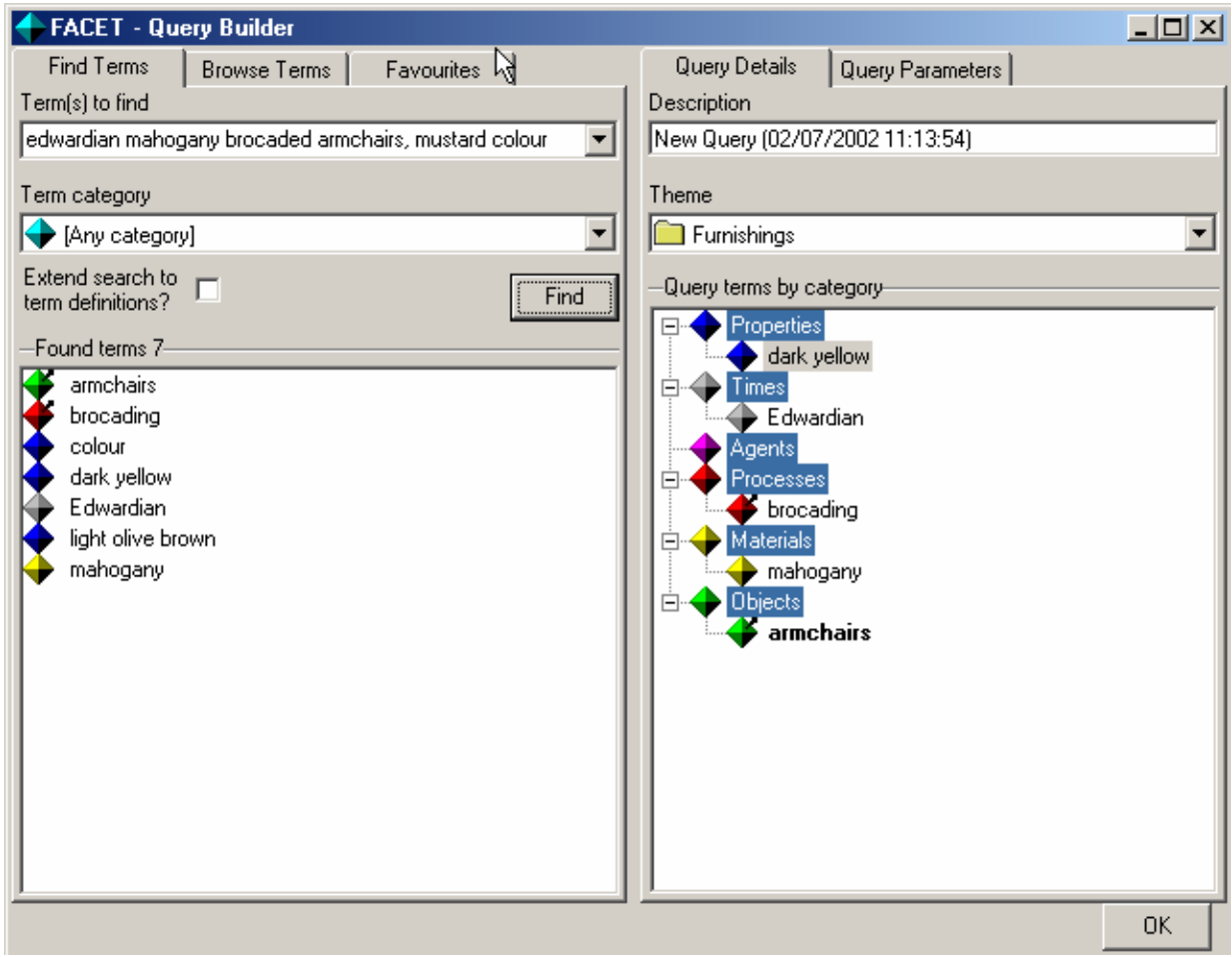
Finding Thesaurus Terms



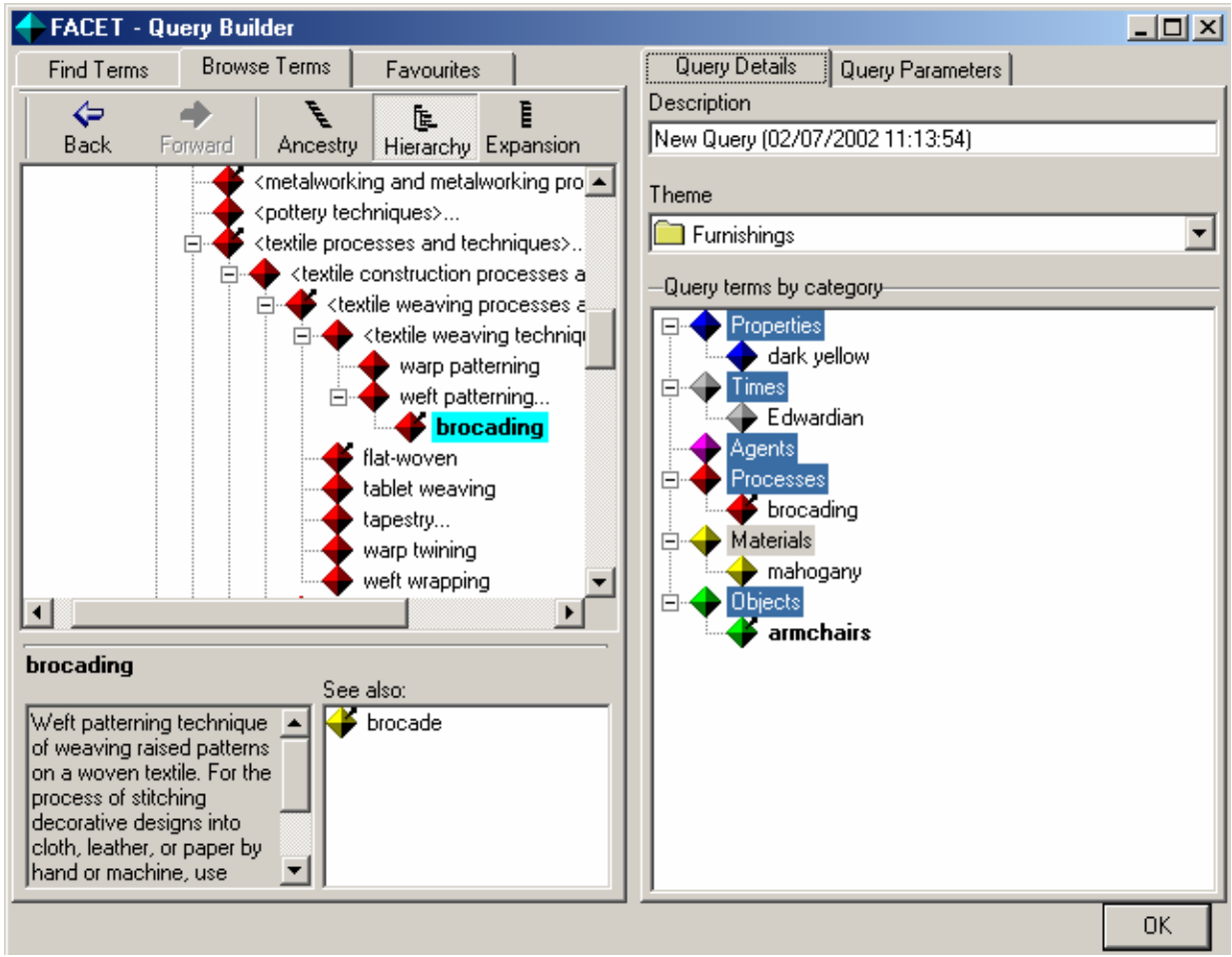
Adding Terms to the Query



Thesaurus Browser in Query Builder



Semantic Expansion Visualisation



Faceted Knowledge Organisation Systems

Faceted classifications based on primary division
into fundamental, high-level categories (facets)

Compound descriptors (multi-concept headings) are synthesised by combination of terms from limited number of fundamental facets

In constructing AAT, adjectival noun phrases very common:

e.g. *painted oak furniture*

“Rather than enumerate the nearly infinite number of object and subject descriptions needed by thesaurus users, the AAT decided to pursue the building blocks of these descriptors in the form of a faceted vocabulary”

(Guide to Indexing and Cataloging with the Art & Architecture Thesaurus)

Compound Descriptors

e.g. *painted oak furniture*

- Multi-concept subject headings allow highly specific descriptions and offer promise of precise queries

- However practical focus has tended to be on cataloguing rather than searching

- Poses problems for recall in retrieval and for browsing.

  Full potential yet to be exploited in retrieval

Matching Problem

"The major problem lies in developing a system whereby individual parts of subject headings containing multiple AAT terms are broken apart, individually exploded hierarchically, and then reintegrated to answer a query with relevance"

(Toni Petersen, AAT Director)

Query: mahogany, dark yellow, brocading, Edwardian, armchair

Descriptor: oak, light yellow, crests, ovals, brocade, Victorian, Carver chair

Potentially extra / missing / partially and non-matching terms

Matching Problem

"The major problem lies in developing a system whereby individual parts of subject headings containing multiple AAT terms are broken apart, individually exploded hierarchically, and then reintegrated to answer a query with relevance"

(Toni Petersen, AAT Director)

Query: mahogany, dark yellow, brocading, Edwardian, armchair

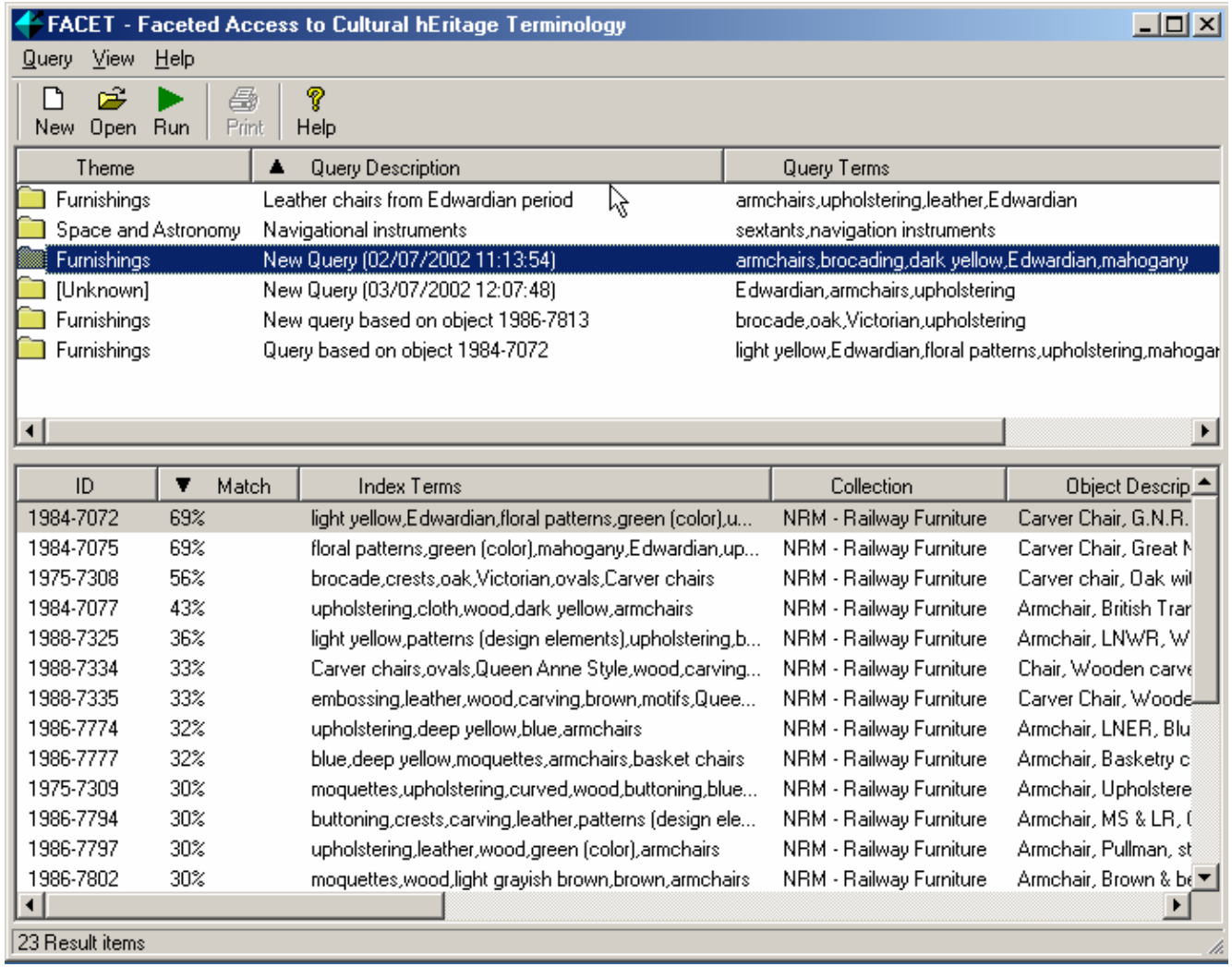
focus term

must match after expansion

Descriptor: oak, light yellow, crests, ovals, brocade, Victorian, Carver chair

Potentially extra / missing / partially and non-matching terms

Facet Queries with Results



Qualitative Evaluation

- **Formative evaluation**

  Analyse at a micro level user interaction in order to illuminate problems

Multiple data sources

transcripts of think-aloud sessions

screen capture movie files

user action logs

observer notes

- **Issues**

  Allocation of search functionality to sub-windows – reduced number sub-windows

  Standalone FACET 2

  Need for integrated Query Builder tool providing more feedback on facet structure

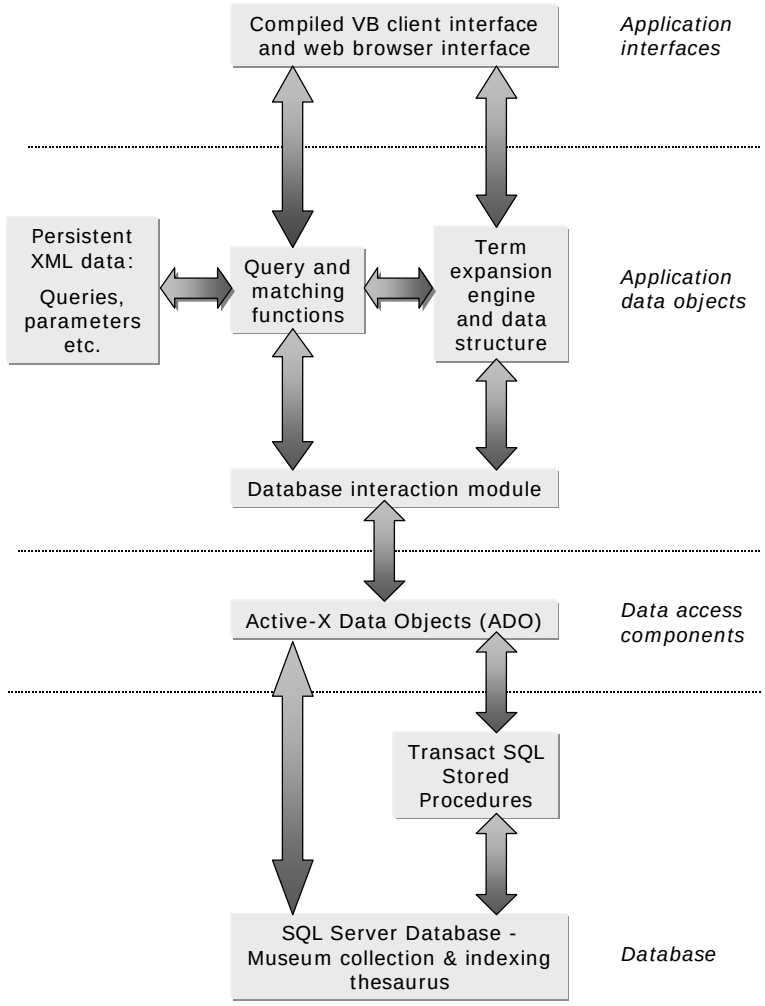
  Introduce notion of focus term in query to avoid distortion on expanding minor terms

  Web Demonstrator

  individual term expansion control

  slider bar inappropriate control

System Architecture



FACET Web Demonstrator

- illustrates thesaurus content and semantic expansion in a fairly realistic Web prototype application

- Intended more as an exploration of FACET research outcomes as dynamically generated Web components than a general interface but suggestive of possible interface components

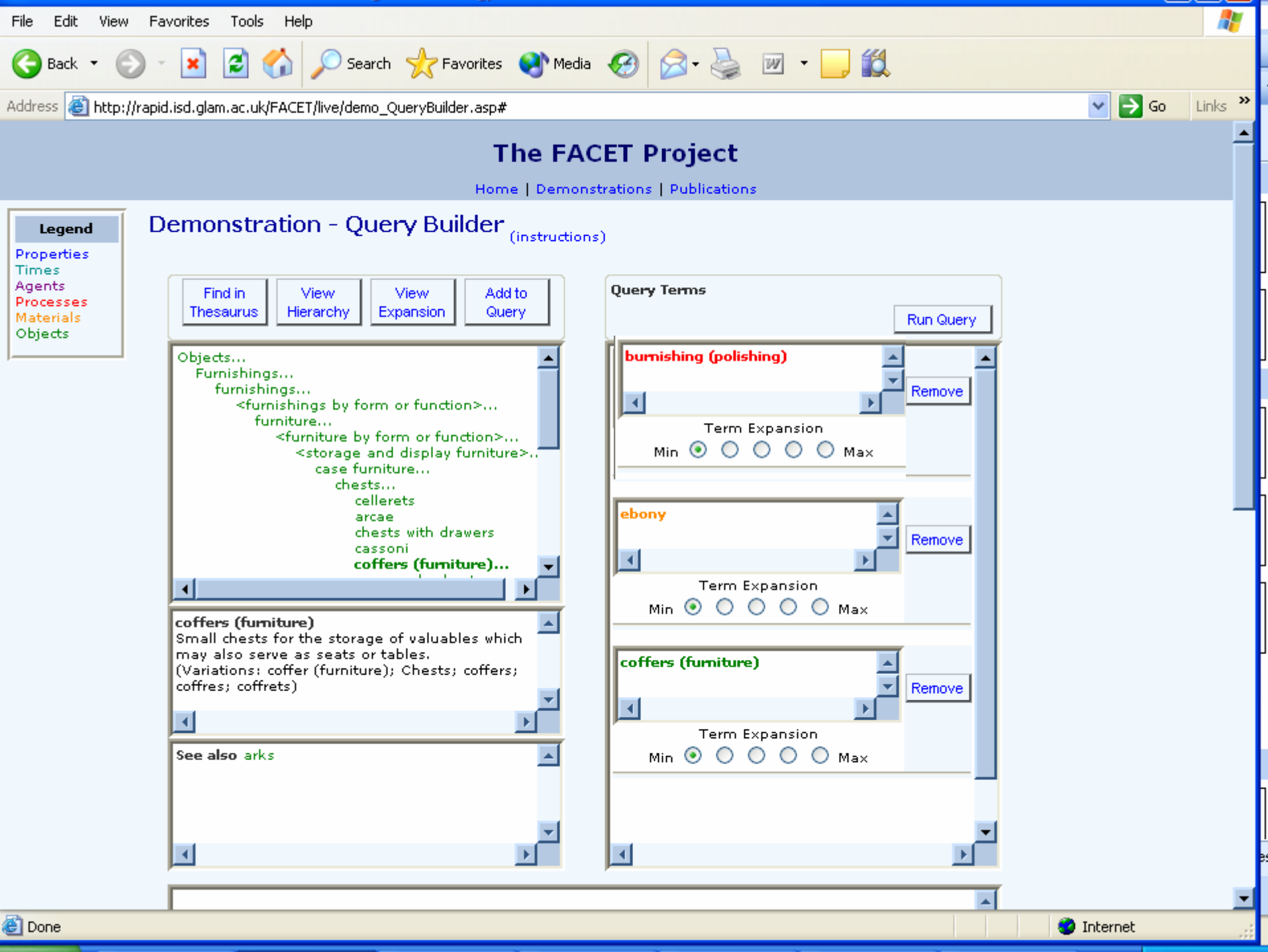
- Not rely on pre-built static HTML pages -
  thesaurus content is generated dynamically

  Both the KOS terminology service (the AAT) and the collections data reside on the same server - in general need not be so

  http://www.comp.glam.ac.uk/~FACET/webdemo/

FACET Web Demonstrator implementation

- Current browser-based interface is an Active Server Pages (ASP) application, using a combination of server-side scripting and compiled components
- Persistence of state information between page requests a problematic issue - HTTP protocol is (by design) stateless
- Solution adopted for current demonstrator involved small 'scriptlet' interface components to communicate with server without causing a browser to refresh the entire page.
- But side effect of introducing some (IE) platform dependence


File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://rapid.isd.glam.ac.uk/FACET/live/demo_QueryBuilder.asp#
Go Links
The FACET Project
Home | Demonstrations | Publications
Legend
Properties
Times
Agents
Processes
Materials
Objects
Demonstration - Query Builder (instructions)
Find in Thesaurus
View Hierarchy
View Expansion
Add to Query
Objects...
Furnishings...
furnishings...
<furnishings by form or function>...
furniture...
<furniture by form or function>...
<storage and display furniture>...
case furniture...
chests...
cellerets
arcae
chests with drawers
cassoni
coffers (furniture)...
coffers (furniture)
Small chests for the storage of valuables which may also serve as seats or tables.
(Variations: coffer (furniture); Chests; coffers; coffres; coffrets)
See also arks
Query Terms
Run Query
burnishing (polishing)
Remove
Term Expansion
Min Max
ebony
Remove
Term Expansion
Min Max
coffers (furniture)
Remove
Term Expansion
Min Max
Done
Internet

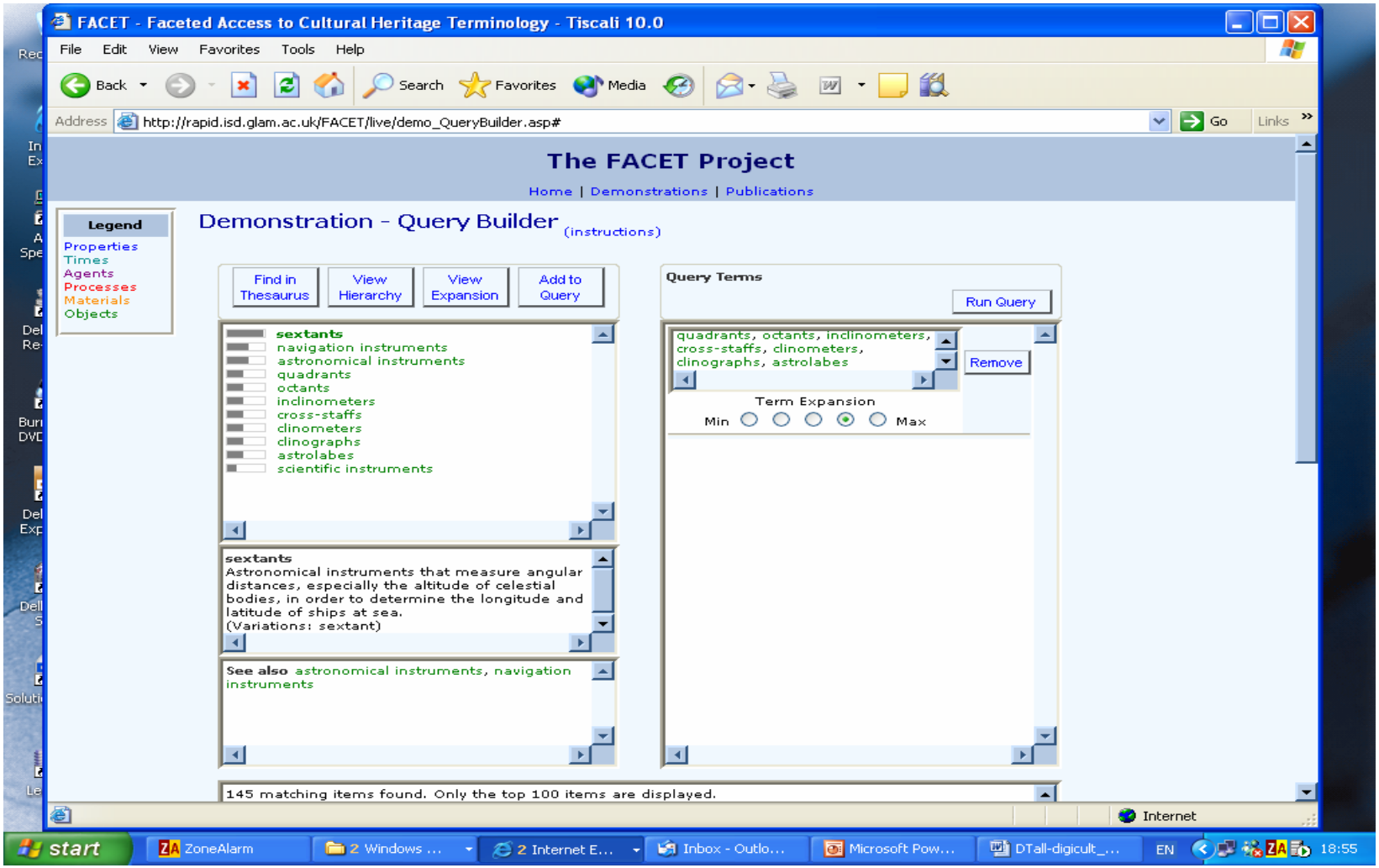
Public Request for Info to Science Museum

- ***“eighteenth century European celestial navigation instruments”***

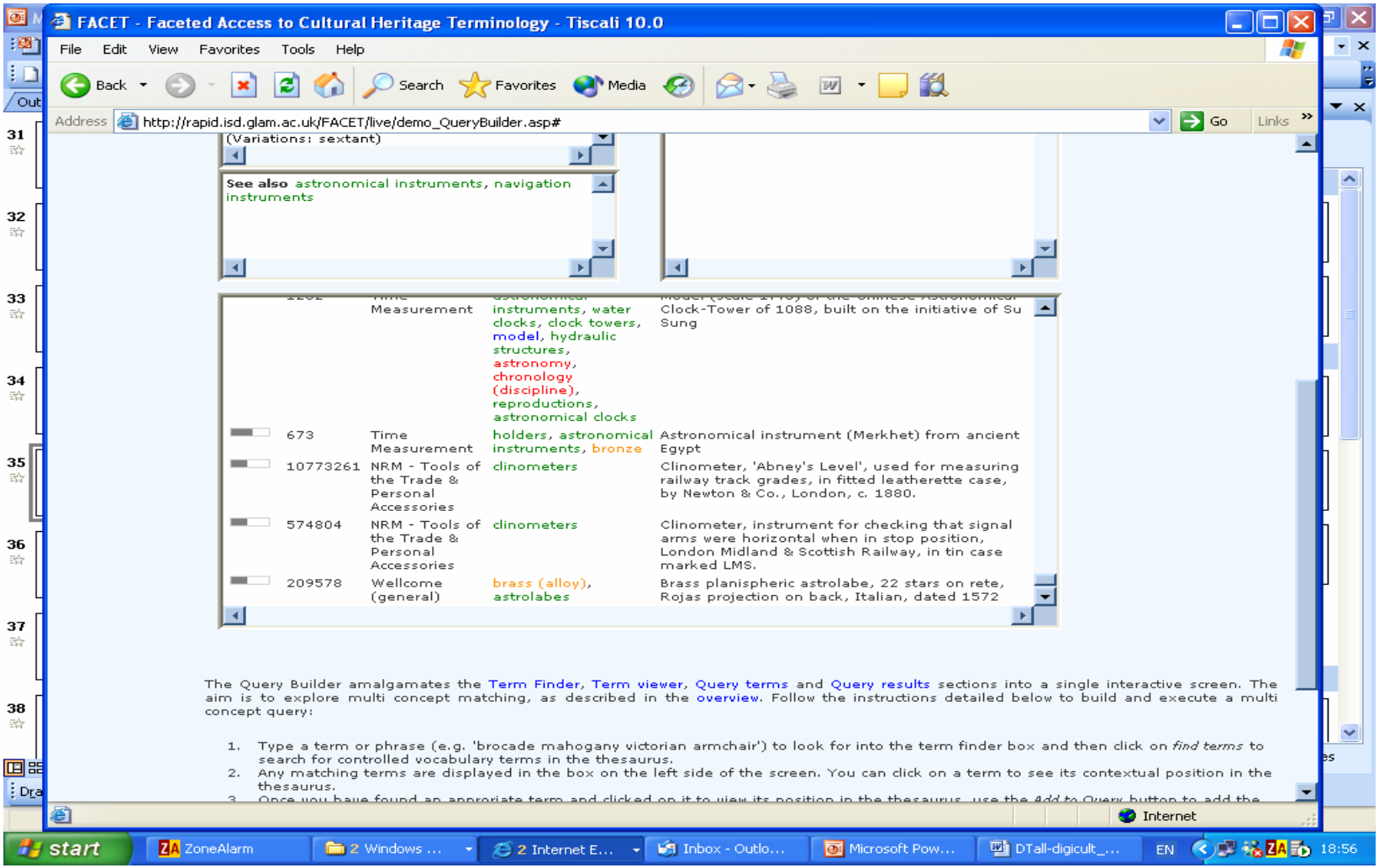
- At the time, this request highlighted difficulties with existing techniques - multiple queries to several fields in database required,

- Semantic expansion on *navigation instruments* short-cuts this process. Items from the collection indexed by *sextants, astrolabes, etc.* now result from a single query.

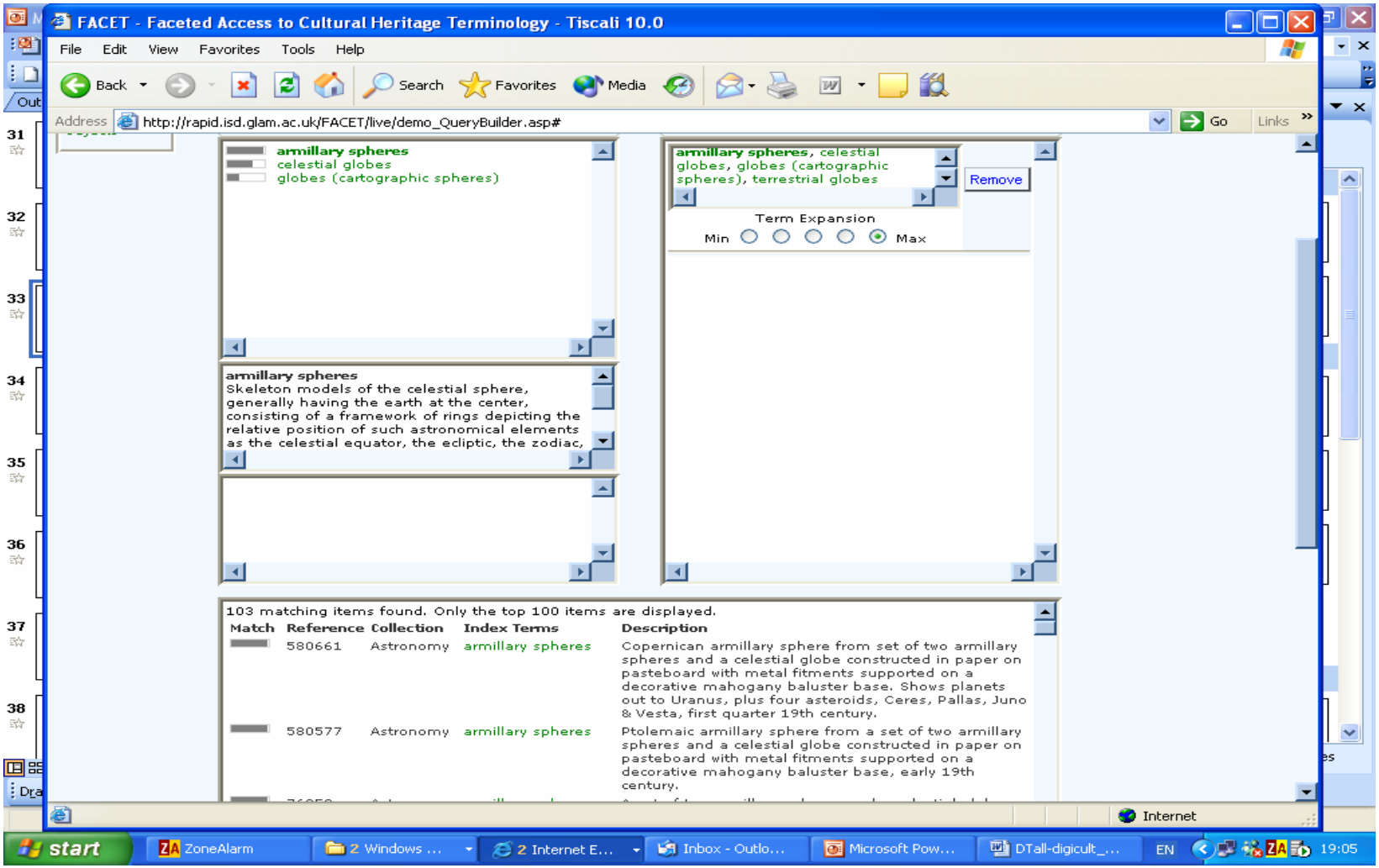
Public Request i



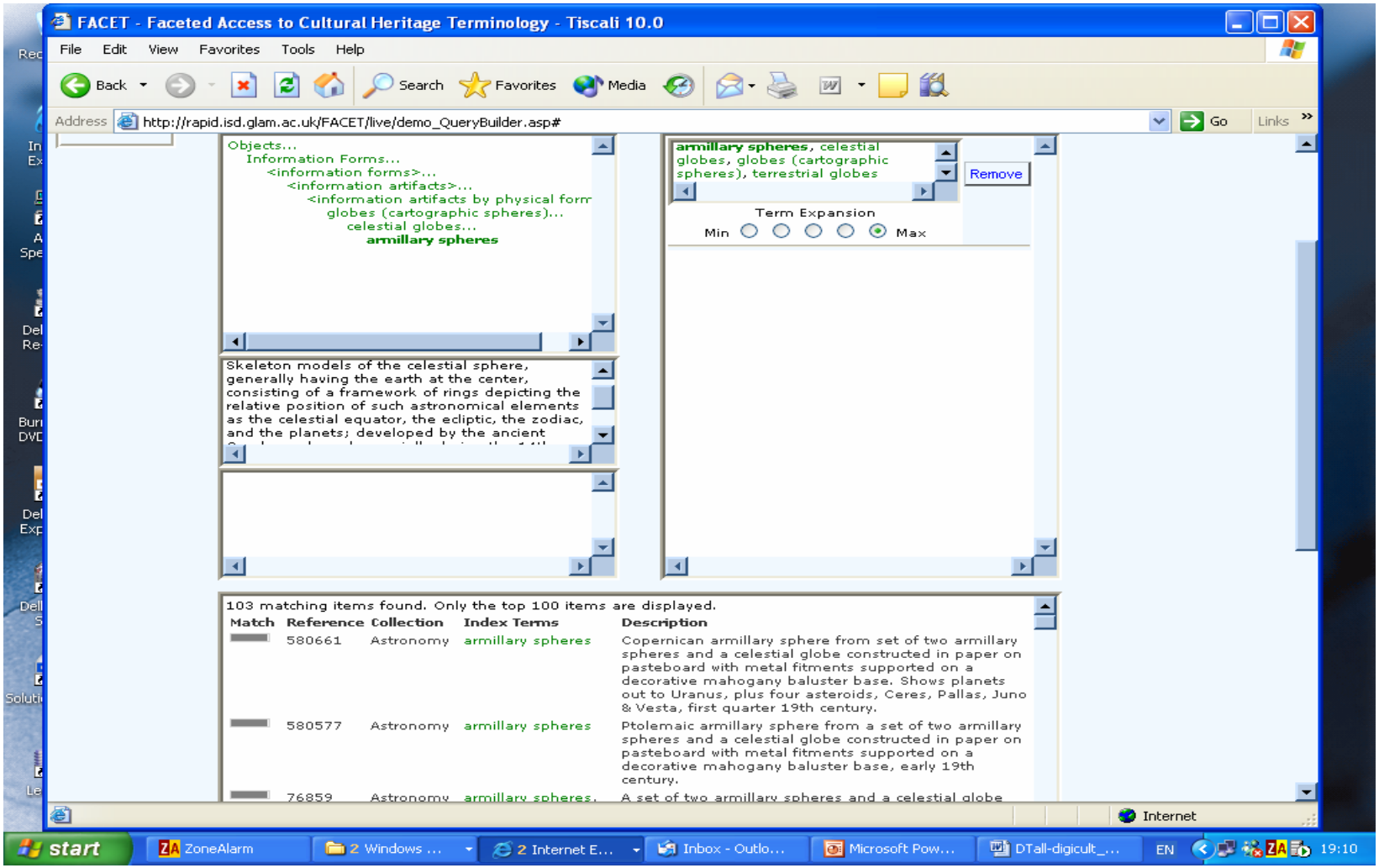
Public Request ii



Public Request iii



Public Request iv



Some lessons learned

- Web demonstrator our first step in exploring issues underlying networked access to KOS
- Results from FACET show that bestmatch (ranked result) approaches can be applied to KOS-based queries via semantic expansion of query terms
- Web interface also showed semantic expansion can also be employed as a browsing tool when wishing to hide some complexity of hierarchical structures

Critical issue - Standards

Ongoing initiatives to revise thesaurus standards

ANSI/NISO Z39.19

BS 5723 and BS 6723 - Dext03

BSI public draft soon, extended scope, interoperability

Thesaurus Representations

RDF - SWAD03; Topic Map - Ligh03; various XML

Possibilities to extend current relationships by specialisation,

enriching standards but maintaining compatibility

KOS Service Protocols - Bind04

NKOS Registry - Vizi01; MEG Registry Project

KOS integration into DL services

from Hill02 Research Agenda KOS/DL

Taxonomy of KOS - KOS types linked to DL service protocols

Registries of KOS and KOS-level metadata to represent them

XML/RDF KOS representations - customisable

Core set of relationship types across all KOS

General KOS service protocol

from which protocols for specific types of KOS can be derived

Robust linking model in which DL entities (collections, objects, and services) can refer to KOS entities (concepts, labels, and relationships)

Visualization tools that fully use and display the rich semantics embedded in KOS

Lessons learned ii

- Critical issue facing KOS in Web is existing standards based in print world and not concerned with data interchange formats.
- Programmatic access requires commonly agreed protocols building on lower-level standards, such as Web services.
- The development of common KOS representation formats and service protocols are closely linked. Progress needs to be made on both dimensions if standards are to be achieved.
- A service protocol should be expressed in terms of a well defined but extensible set of KOS data elements and relationships, with the relationship type a parameter to the protocol commands. This would allow the specialisation of the current thesaurus relationships.

Lessons learned iii

- Trend towards service oriented architectures brings opportunity of clearer separation of interface components from underlying data sources, via use of appropriate (Web) services

- Basing distributed protocol services on atomic elements of thesaurus data structures and relationships would limit possible interfaces (too many protocol requests to server)

- Web interfaces offering advanced thesaurus services require protocols which group primitive thesaurus data elements (via their relationships) into *composites*, to achieve reasonable response rate.

How far to formalise KOS?

Thesaurus a long-lived, pragmatic and useful tool

- includes semantics, domain lexicon (UF/ALTs, Scope Notes)
- cost-effective granularity of relationships for many search apps
  - where results are based on probable relevance judgements

Cost/benefit issues in extent of KOS formalisation

- Application domain dependent level of precision in concept use
- Indexer - Searcher variation in applying concepts
- Formalisation depends on how applications process concepts

Existing KOS already have rich resources to offer,

- not withstanding future semantic web developments
- which will tend to be more resource intensive

Some current work on semantic DL approaches

- FACET → WebFACET → **InterFACET**

- Next phase of work looks at common KOS representation formats and API protocols - making content available via programmatic interfaces.
- portable, platform neutral, open-source code

- One focus: ***semantic expansion as a service***
  - possible KOS protocol element?

yields

- different configurations of KOS displays by single function call
- novel alternative interfaces, such as navigation via semantic expansion
- Automatic expansion of query terms for various ranked result (best match) query services
- Term suggestion facilities to assist in document indexing applications

Current work examples

- Import (currently extract just semantic-structural parts)
  AAT-REC format files (Getty format)
  MultiTes CSV export format files
  SKOS-Core format files

  AAT, ADL Feature Type Thesaurus
  EIONET-GEMET thesaurus (SKOS-Core format)

- into expansion module and perform semantic expansion

- Export to SKOS-Core (approx) format files

Current work examples – AAT SKOS RDF export

```
<skos:Concept rdf:about="201687">
<skos:prefLabel>sextants</skos:prefLabel>
<skos:related rdf:resource="256965" />
<skos:altLabel>sextant</skos:altLabel>
<skos:broader rdf:resource="196710" />
<skos:related rdf:resource="24497" />

</skos:Concept>
<skos:Concept rdf:about="195790">
<skos:prefLabel>astrolabes</skos:prefLabel>
<skos:altLabel>astrolabe</skos:altLabel>
<skos:broader rdf:resource="196710" />
<skos:related rdf:resource="24497" />
</skos:Concept>
```

```
<skos:Concept rdf:about="24497">
<skos:prefLabel>astronomical
instruments</skos:prefLabel>
<skos:related rdf:resource="251656" />
<skos:related rdf:resource="201687" />
<skos:related rdf:resource="54534" />
<skos:related rdf:resource="195790" />
<skos:related rdf:resource="196040" />
<skos:related rdf:resource="199777" />
<skos:altLabel>astronomical
tools</skos:altLabel>
…
</skos:altLabel>
<skos:broader rdf:resource="122283" />
<skos:related rdf:resource="25789" />
</skos:Concept>
```

Current work examples – sextants expanded

1.000		sextants
0.697	\BT	<vertical angle measuring devices>
0.611	\RT	astronomical instruments
0.611	\RT	navigational instruments
0.522	\RT\UF	instruments, astronomical
0.522	\RT\UF	astronomical instrument
...		
0.477	\BT\NT	clinographs
0.477	\BT\NT	clinometers
0.477	\BT\NT	quadrants
0.477	\BT\NT	cross-staffs
0.477	\BT\NT	octants
0.477	\BT\NT	inclinometers
0.477	\BT\NT	astrolabes
0.390	\BT\NT\UF	cross-staff
0.390	\BT\NT\UF	inclinometer
0.390	\BT\NT\UF	staffs, Jacob's
0.390	\BT\NT\UF	Jacob's staffs

Current work – GEMET example expanded

1.000		landslide
0.591	\BT	geomorphic process
0.357	\BT\NT	avalanche
0.357	\BT\NT	erosion
0.169	\BT\RT	geological process
0.169	\BT\BT	land

Idea is a semantic expansion service
which an application might use for

- *browsing display*
- *query expansion*
- *indexing suggestions, etc*

NKOS Workshop at ECDL 2004

- NKOS Workshop –
  User-centred approaches to NKOS

  ECDL 2004, Bath, UK, 16 September
  see http://www2.db.dk/nkos-workshop/

- Selected papers from the NKOS workshop will be considered for forthcoming special issues of journal NRHM

Contact Information

Doug Tudhope
School of Computing
University of Glamorgan
Pontypridd CF37 1DL
Wales, UK

dstudhope@glam.ac.uk
http://www.comp.glam.ac.uk/pages/staff/dstudhope

References

Aitchison J., Gilchrist A., Bawden D. 2000. Thesaurus construction and use: a practical manual (4th edition). London: ASLIB.

Binding C., Tudhope D. 2004. KOS at your Service: Programmatic Access to Knowledge Organisation Systems. JoDI 4(4), http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Binding/

Blocks D., Binding C., Cunliffe C., Tudhope D. 2002. Qualitative evaluation of a thesaurus-based retrieval system. Proc. ECDL 2002, 346-361. LNCS. © Springer-Verlag. http://www.glam.ac.uk/soc/research/hypermedia/publications/presentationdocs/ecdl.pdf

Dextre Clarke S. 2003. BS 8723 : a new British Standard for structured vocabularies. http://www.glam.ac.uk/soc/research/hypermedia/NKOS-workshop%20Folder/dextre_clarke.ppt

FACET Project. http://www.comp.glam.ac.uk/~FACET/

FACET Web demonstrator. http://www.comp.glam.ac.uk/~FACET/webdemo/

Hill et al. 2002. Integration of Knowledge Organization Systems into Digital Library Architectures. ASIST SigCR - http://www.lub.lu.se/SEMKOS/docs/Hill_KOSpaper7-2-final.doc

Hodge Gail, 2000. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. CLIR Pub91. April 2000. http://www.clir.org/pubs/abstract/pub91abst.html

Jacob Elin. 2003. Ontologies and the Semantic Web. ASIST Bulletin, April/May 2003, Special Issue on Semantic Web. http://www.asis.org/Bulletin/Apr-03/BulletinAprMay03.pdf

Koch T. Activities to advance the powerful use of vocabularies in the digital environment - Structured overview. http://www.lub.lu.se/~traugott/drafts/seattlespec-vocab.html

Light R. 2003. XML (and Topic Maps). http://www.richardlight.org.uk/thesauri/thesauri.htm

McGuinness D. 2002. Ontologies Come of Age. In: (Fensel et al eds.) Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press.

MultiTes 2003. Conference on Thesauri and Taxonomies http://www.multites.com/conference03.htm

References ctd.

NKOS: Networked Knowledge Organization Systems/Services, http://nkos.slis.kent.edu/

NKOS 2003. Workshop ECDL. http://www.glam.ac.uk/soc/research/hypermedia/NKOS-Workshop.php

NKOS 2004. New Applications of Knowledge Organization Systems. NKOS Special Issue, JoDI. http://jodi.ecs.soton.ac.uk/?vol=4&iss=4

Noy N., McGuinness D. Ontology Development 101: A Guide to Creating Your First Ontology. http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

Soergel D. The representation of Knowledge Organization Structure (KOS) data: a multiplicity of standards. http://www.glam.ac.uk/soc/research/hypermedia/publications/SoergelNKOS2001KOSStandards

SWAD-Europe Thesaurus Activity. http://www.w3.org/2001/sw/Europe/reports/thes/

Tudhope D., Binding C., Blocks D., Cunliffe D. 2002. Compound Descriptors in Context: A Matching Function for Classifications and Thesauri. Proc. JCDL 2002, 84-93.

http://www.glam.ac.uk/soc/research/hypermedia/publications/jcdl02.pdf

Vizine-Goetz D. 2001. NKOS Registry - draft proposal for KOS-level metadata. http://staff.oclc.org/~vizine/NKOS/Thesaurus_Registry_version3_rev.htm

Extract from Collated Transcript

Notation indicates data source

The user searches the thesaurus for "text":
52:08.1Thesaurus form: Click "Find now"Text
The user then adds "text" to the query. After this, he looks at the related terms, and drags the (only) related term ("words") also into the query. The user executes the query.
53:06.4Query form: QueryStartStart Query: Text, Words
53:12.0Query form: Results: 3
Three records come up as a result:
IDMatchCollectionDescription
The user looks at the first record. The indexing terms are: cast iron, embossing, inlays, lettering (layout features), seating, wood.
53:57.0Catalogue record: Activate window
Participant: Right, okay, so you've got the words "embossing". Ah, okay. That's just a different tense, isn't it? … Participant: So I can now try that and see if it will get me anything more.